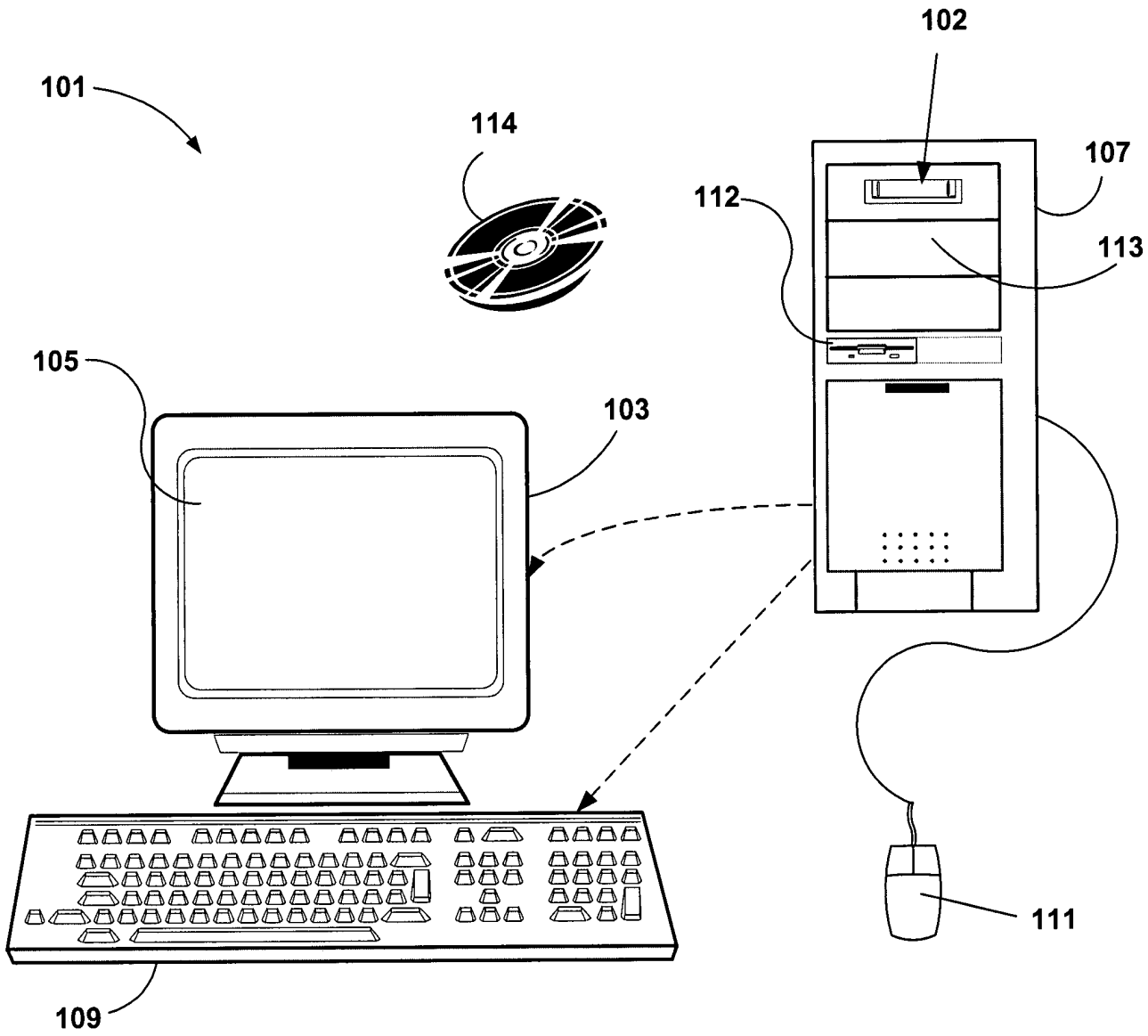


Figure 1

Attorney Docket No. 21525-0004

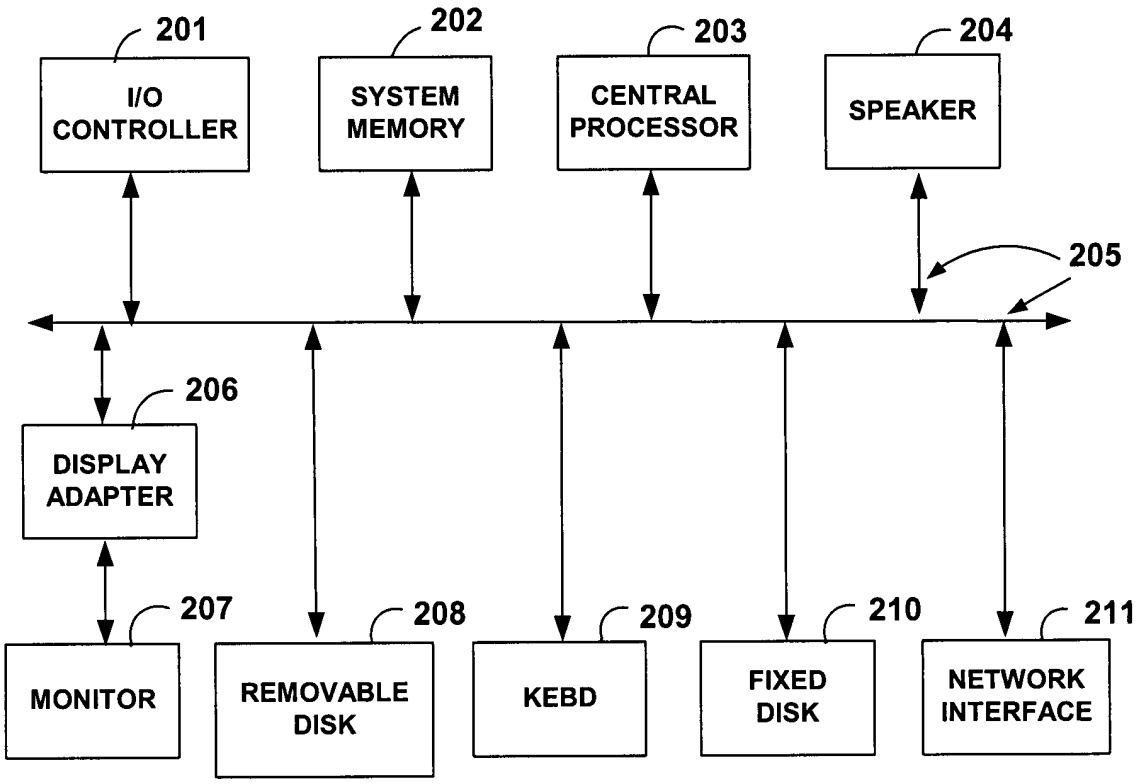


Figure 2

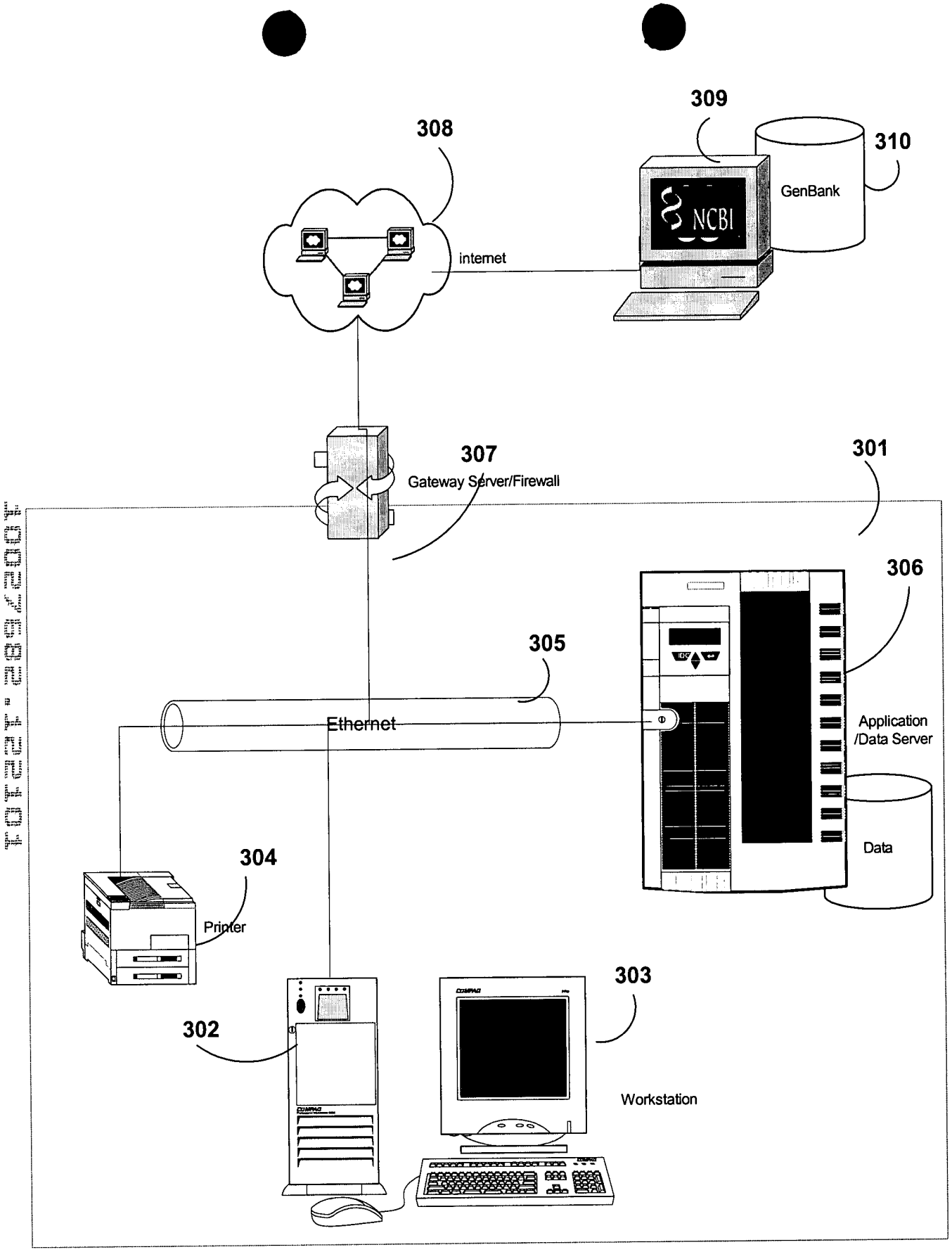

309
308
310
GenBank
NCBI
internet
307
Gateway Server/Firewall
301
306
305
Ethernet
Application
/Data Server
Data
304
Printer
303
302
Workstation

FIGURE 3

Docket No. 4505-4

Sequence Collection

- Sequences and annotations collect from a variety of sources.

Sequence Analysis

- Aligned to the draft assembly of the genome.
- Low quality bases trimmed from ESTs.
- Polyadenylation sites detected.

Seed Cluster Creation

- Seed clusters created for each UniGene Cluster.
- Additional seed clusters are created with potential full-length mRNAs not in UniGene.

Genomic Based Subclustering

- Seed clusters are subclustered based on the contig.

Sequence Based Subclustering

- Sequences are subclustered in transcriptome space.

Orientation Based Subclustering

- EST read directions, CDS annotations, consensus splice sites, and polyadenylation sites are used to predict the orientation of the subcluster. The subcluster is resubclustered if substantial conflicts exist.

Picking Probe Selection Regions

- For each subcluster, regions for probe selection are selected from either the consensus or exemplar sequences.
- Multiple regions may be choosen due to alternative polyadenylation sites.

Prioritizing Sequence Selection Regions

- Regions for probe selection are prioritized based on the quality of sequence and annotation.

FIGURE 4

Page 23 of 23 - 09/29/2004 4:09:25 PM

FIGURE 5

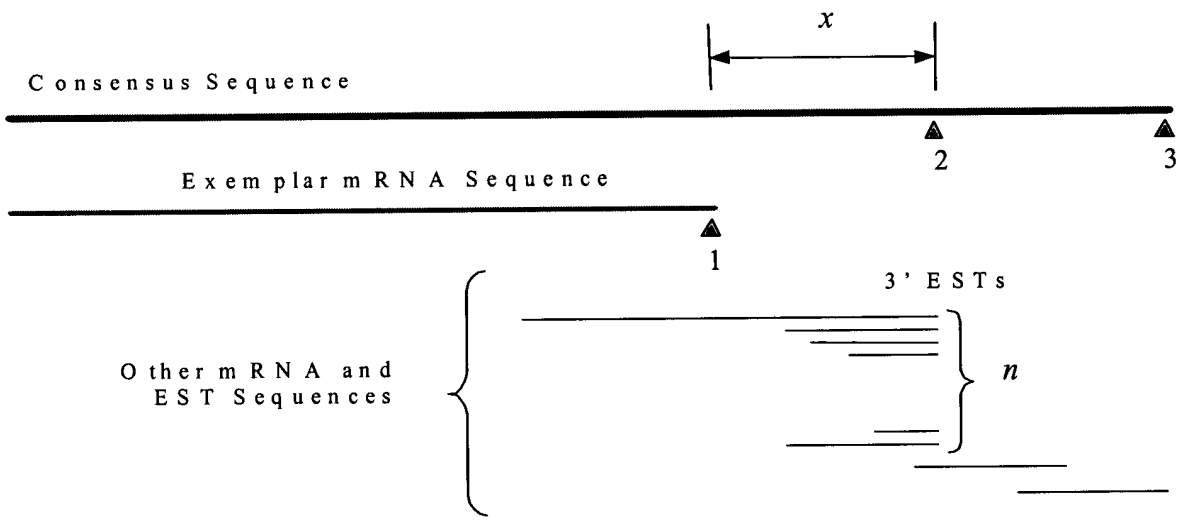
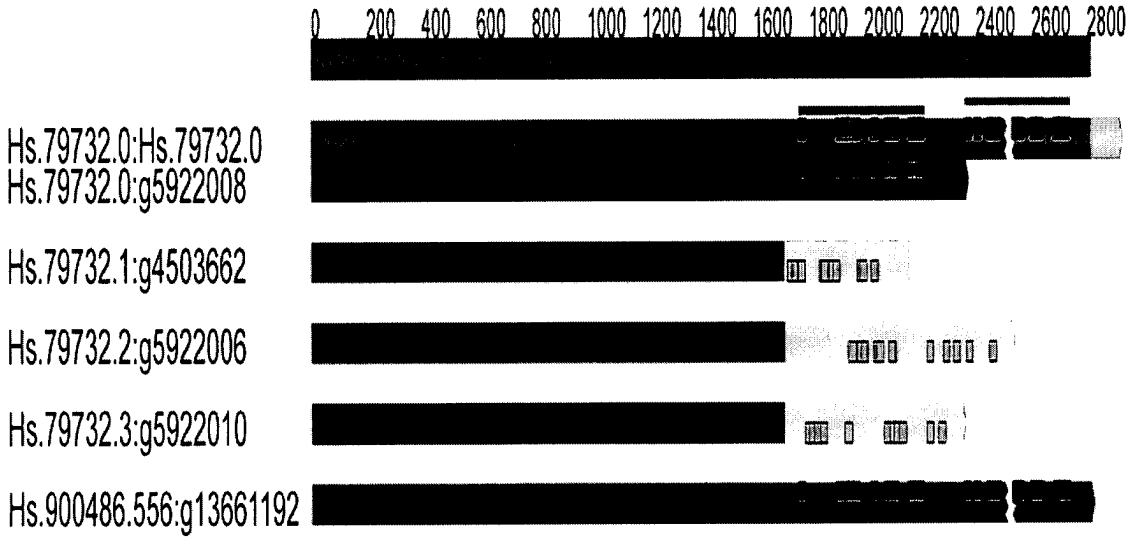

x
Consensus Sequence
2
3
Exemplar mRNA Sequence
1
3' ESTs
Other mRNA and
EST Sequences
n

FIGURE 6



EDS3 6802 - 4202 4804

Figure 7